



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Simple dissociations for a higher-powered neuropsychology

Citation for published version:

McIntosh, R 2018, 'Simple dissociations for a higher-powered neuropsychology' Cortex, vol. 103, pp. 256-265. DOI: [10.1016/j.cortex.2018.03.015](https://doi.org/10.1016/j.cortex.2018.03.015)

Digital Object Identifier (DOI):

[10.1016/j.cortex.2018.03.015](https://doi.org/10.1016/j.cortex.2018.03.015)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Cortex

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Simple dissociations for a higher-powered neuropsychology

Robert D McIntosh

Human Cognitive Neuroscience, Psychology, University of Edinburgh, UK.

7 George Square, Edinburgh, EH8 9JZ

Tel: +44 131 6503444

Fax: +44 131 6503461

r.d.mcintosh@ed.ac.uk

Open-data and materials: R code for the empirical simulations, and the full data reported, are uploaded with this submission.

Abstract

Dissociations in cognitive neuropsychology are often investigated at the level of the single-case, and formal criteria exist for the detection of dissociations, and their sub-classification into ‘classical’ and ‘strong’ types. These criteria require a patient to show a frank deficit on one task (for a classical dissociation) or both tasks (for a strong dissociation), and a significantly extreme difference between tasks. I propose that only the significant between-task difference is logically necessary, and that if this simple criterion is met, the patient should be said to show a dissociation. Using Monte Carlo simulations, I show that this simplification increases the power to detect dissociations across a range of practically-relevant conditions, whilst retaining excellent control over Type I error. Additional testing for frank deficits on each task provides further qualifying information, but using these test outcomes to categorise dissociations as classical or strong may be too uncertain to guide theoretical inferences reliably. I suggest that we might instead characterise the strength of the dissociation using a continuous index, such as the effect size of the between-task difference.

Keywords: Single-case; double dissociation; statistical methods.

1. Introduction

Dissociations are a core concept in cognitive neuropsychology. If two putative mental functions can be disturbed, each without comparable consequences for the other, we may hypothesise some degree of independence between them (i.e. *double dissociation*; Teuber, 1955). At a practical level, dissociations and double-dissociations are empirical observations. These observations can support theoretical inferences about the functional architecture of cognition, provided that certain assumptions are met. The inferences that dissociations can license, and the plausibility of the assumptions required for them to hold, have been discussed and debated elsewhere (Caramazza, 1986; Caramazza & McCloskey, 1988; Coltheart, 2001, 2017; Dunn & Kirsner, 2003; Ellis & Young, 1988; Patterson & Plaut, 2009; Shallice, 1979, 1988, 2015). In this paper, I focus on a more limited, practical question: if neuropsychologists *are* interested in finding dissociations, what operational criteria should they use?

Dissociations are often investigated at the level of the individual ‘case’, in isolation or as part of a larger series, and a range of statistical tests have been developed to compare single cases against matched control samples, as a proxy for the patient’s pre-morbid abilities (see McIntosh & Brooks, 2011, for a brief overview). These tests estimate how unlikely it would be to find performance more extreme than that of the single case, if testing a person from the healthy population. Case-control comparisons are inherently low-powered, a problem exacerbated by the common use of very small control samples ($n \leq 10$) (Crawford & Garthwaite, 2006a, 2006b; Crawford, Garthwaite, & Gray, 2003). The saving grace of cognitive neuropsychology is that the effects of brain damage are often severe, and thus visible to low-powered statistical microscopes. Without such large effects, single-case cognitive neuropsychology would barely be possible at all. Nonetheless, as Crawford, Garthwaite, & Ryan (2011) have argued, “*anything that can increase statistical power to detect deficits or dissociations should be encouraged (provided that it does not achieve this at the cost of failing to control the Type I error rate)*” (p. 1167).

I propose that a simple change could be made to current criteria that would increase the power to detect dissociations, under a range of practically-useful conditions, whilst retaining appropriate control over Type I error rate. I will explore these intuitions empirically, using Monte Carlo simulations. Before doing so, I provide a brief background on the operational definition of deficits and dissociations in cognitive neuropsychology, and recent

statistical developments. Readers familiar with these ideas may wish to skip directly to the methodological proposal (Section 3).

2. Deficits and dissociations

The classical concepts of deficit and dissociation are easy to grasp: acquired brain-damage has impaired some aspect of a person's mental abilities (deficit), which is at odds with the preservation of some other aspect (dissociation). The modern framework for single-case dissociations, and the assumptions required for inferences on mental structure, was laid out most comprehensively by Shallice (1988). Amongst the contributions of his landmark book was the delineation of three forms of dissociation, of differing degrees of inferential strength (Shallice, 1988, Chapter 10).¹ The most powerful, *classical dissociation*, corresponds roughly to the stark contrast typical of the classical literature, with performance on one task (X) grossly impaired whilst another task (Y) is performed normally. Of intermediate strength, and thus perhaps misleadingly named, was the *strong dissociation*, defined by a striking discrepancy between the performance of tasks X and Y, but with deficits on both tasks. A more minor, *trend dissociation* was also suggested, in which two tasks are performed at different levels, yet without the striking discrepancy that characterises a strong or classical dissociation. Shallice noted that the distinction between the strong and trend forms, might be somewhat loose in practice, in part because "*it would be very difficult to calculate for any particular pair of tests what spread of performance would be expected to arise in the normal population*" (p240).

Fifteen years later, Crawford and colleagues (2003) enumerated three key problems with these criteria, as applied within the research field. First, there was yet no widely-agreed statistical test for a single-case deficit. A deficit might be demonstrated by reference to a standardised cut-off, by quantitative comparison (e.g. z-score) with a small to modestly-sized control group ($n = 5-15$), or just inferred from a qualitative contrast with one or more 'representative' controls. The second problem was that the definition of a classical dissociation requires preserved performance on one task, but this cannot be confirmed by conventional means, because it requires us to 'prove' the null hypothesis of no deficit, when at best we can fail to reject it. Third, and most crucially, classical and strong dissociations imply an abnormal discrepancy in the degree of impairment between tasks, but no validated

¹ Shallice accompanied these definitions by a careful consideration of the assumptions required, and some recommendations for maximising the likelihood of valid inferences from dissociations. These included the use of multiple tasks to converge on the functions of interest, because, "*Inferences from individual patients that are based on only a single test findings are, in my opinion, highly suspect.*" (Shallice, 1988, p231). For simplicity of the present treatment, however, we consider the scenario of a single task X and single task Y.

method was available for testing this directly (see Shallice's comment, above). A positive test of the all-important *dissociation* between tasks was lacking.

Crawford and colleagues were able to suggest solutions for the first and third of these problems. They noted that estimating the rarity of a patient's score by reference to the z-distribution assumes that the control mean and standard deviation are population parameters, when very often they are biased estimates from a restricted, and often small, sample. In restricted samples ($n < 50$), the likelihood of extreme z-scores will be underestimated, promoting false positive findings of deficits (high Type I error rates). Their solution was instead to base the test of deficit on the t-distribution, taking account the size of the control sample. Crawford and Howell (1998) developed a modified t-test for case-control comparisons, which constrains Type I error rate robustly, even with very small control samples ($n < 10$), and which is also reasonably tolerant of departures from normality (Crawford, Garthwaite, Azzalini, Howell, & Laws, 2006).

Similarly, Crawford, Howell, and Garthwaite (1998) modified the *paired* t-test, to provide a parametric method for comparing the difference between a patient's performances on two tasks against the distribution of paired differences amongst controls. They devised an *Unstandardised Difference Test*, appropriate if the two tasks are on a common scale and performed with similar variances by controls, and a *Revised Standardised Difference Test* for the more usual scenario of two differently-scaled tasks, which standardises the scores on each task before the differences are assessed (Crawford & Garthwaite, 2005). Crawford and colleagues subsequently developed Bayesian counterparts for these tests. The Bayesian Test of Deficit gives similar outcomes to the modified t-test; the Bayesian Unstandardised Difference Test gives similar outcomes to the Unstandardised Difference Test; but the Bayesian Standardised Difference Test outperforms the Revised Standardised Difference Test (Crawford & Garthwaite, 2007b; Crawford et al., 2011). These Bayesian tests have subsequently been extended further to allow for the inclusion of covariates (Crawford et al., 2010). Most importantly, for present purposes, the establishment of robust case-control tests of deficit and between-task difference enabled Crawford et al. (2003) to define formal criteria for Shallice's categories of *classical* and *strong* dissociation, the *trend* dissociation being dropped entirely (Table 1).

Dissociation	Compound criteria
Classical	1. Deficit on task X: Patient's score significantly lower than controls ($p < .05$, <i>one-tailed</i>). 2. No deficit on task Y: Patient's score not significantly lower than controls ($p > .05$, <i>one-tailed</i>). 3. Between-task difference: Patient's discrepancy significantly more extreme than controls ($p < .05$, <i>two-tailed</i>)
Strong	1. Deficit on task X: Patient's score significantly lower than controls ($p < .05$, <i>one-tailed</i>). 2. Deficit on task Y: Patient's score significantly lower than controls ($p < .05$, <i>one-tailed</i>). 3. Between-task difference: Patient's discrepancy significantly more extreme than controls ($p < .05$, <i>two-tailed</i>)

Table 1. Compound criteria for dissociations (classical and strong), as developed by Crawford, Garthwaite, & Gray (2003; revised by Crawford & Garthwaite, 2005, 2007a). Different tests are available both for within-task deficits and between-task differences, appropriate to different circumstances. The present focus is not on which tests are used, but on how the test results are combined by the compound criteria.

As summarised in Table 1, a classical dissociation is given by a significant deficit on Task X, no significant deficit on Task Y, and a significant between-task difference, while a strong dissociation is given by significant deficits on both tasks, and a significant between-task difference. I refer to these as the *compound criteria*, because they depend upon combined outcomes, across two tests of deficit and one test of difference. The precise tests of deficit and difference used, whether frequentist or Bayesian, unstandardized or standardised, and one- or two-tailed, may vary with circumstances; the critical point is the way in which the test results are combined. A strong dissociation is founded on positive findings across all three tests, whilst a classical dissociation is distinguished by one failure to reject the null hypothesis of no deficit (on Task Y). Considering the inherently low-power of case-control comparisons, it is thus unsurprising that a true strong dissociation will often be misclassified as classical, while the reverse mistake will rarely be made (Crawford & Garthwaite, 2006a). Given this common misclassification, Crawford and Garthwaite concluded that the strong/classical distinction is of limited value, and recommended that the classical dissociation should be re-named as “a dissociation (putatively classical)” (p. 2257). For brevity here, I use the more straightforward term, classical dissociation.

3. A methodological proposal

By the compound criteria (Table 1), a dissociation between two tasks (X and Y) requires a patient to be deficient on at least one task, and additionally to have an abnormally large difference between them (Crawford et al., 2003). Crawford & Garthwaite (2007, p362) identified the abnormal difference as the “most important” component. I propose that we can go further and say that an abnormal between-task difference is the *only* component necessary to the concept of dissociation; if this *simple criterion* is met, the patient should be said to show a dissociation between tasks. Additional testing for deficits in the individual tasks may qualify the dissociation further, but is strictly irrelevant to the question of whether a dissociation exists.

We can prime the logic of this proposal using the model example in Figure 1. The unfilled circles are identical in all plots, and show the scores of ten control participants on two tasks (X and Y, correlated at $r = .5$), alongside the distribution of paired XY differences. For convenience, the two tasks are on a common scale, and performed with comparable variance by controls, which allows us to use the *Unstandardised Difference Test* (UDT) to assess between-task differences. For added simplicity, the (unstandardized) mean control score on each task happens to be 0, with an SD of 1. The dotted lines indicate the one-tailed threshold for deficit on each task by the modified t-test ($p < .05$), and the two tailed threshold for abnormal XY difference by the UDT ($p < .05$). The filled circles show a single-case, to whom the healthy controls are matched, and depict different scenarios in Figures 1a and 1b.

Figure 1a shows one hypothetical pattern of *pre-morbid* performance for our single case, and then how she would present following brain damage that selectively decreased her ability to perform task X, by 2SD. Given that she exceeds the cut-off on the test of deficit for task X, and exceeds the cut-off for abnormal difference, then either the compound or simple criteria would correctly detect the dissociation. Next, Figure 1b shows a scenario in which our patient had a generally higher level of pre-morbid ability, but sustained exactly the same selective 2SD impairment on task X. Given her greater premorbid capacity for task X, this impairment is insufficient to push our patient out of the normal range, and the modified t-test fails to find a deficit. But the outcome of the difference test is identical to that in the first scenario (Figure 1a), because the same discrepancy has arisen between tasks. The compound

criteria would conclude that there is no dissociation, because the patient is deficient on neither task, but the simple criterion would correctly detect the dissociation.

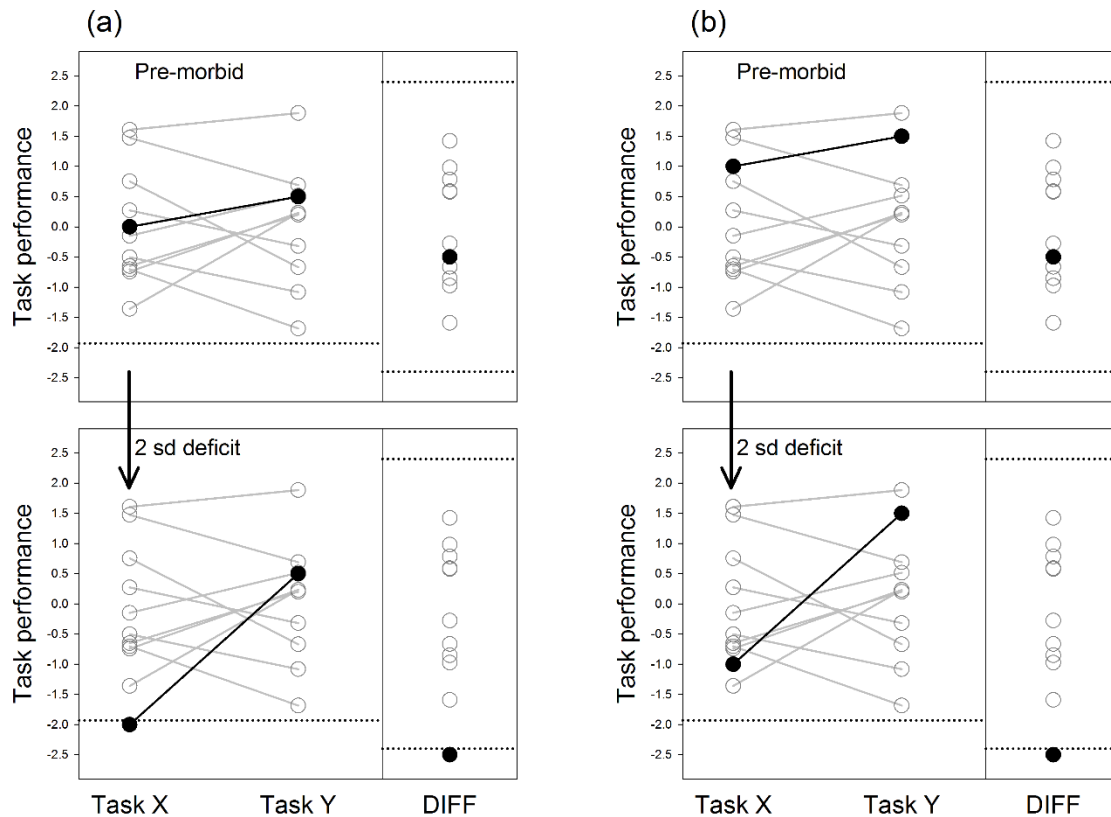


Figure 1. Hypothetical data from a single case (filled circles) and ten matched controls (unfilled circles), on two tasks (X and Y , correlated at $r = .5$). The dotted lines indicate one-tailed cut-offs for deficit on each task, and the two-tailed cut-offs for between-task difference ($p < .05$). Each side of the figure depicts a different scenario. **(a)** The patient's hypothetical pre-morbid ability is shown, and is roughly in the middle of the normal range. Brain-damage induces a selective 2SD decrease in ability to perform task X , and the patient exceeds cut-offs for deficit and difference, showing a dissociation according to the simple criterion or the compound criteria. **(b)** A different hypothetical level of pre-morbid ability is shown, in the upper half of the normal range. Brain-damage induces a selective 2SD decrease in task X , and the patient exceeds the cut-off for difference, but not deficit, showing a dissociation according to the simple criterion, but not the compound criteria. In scenarios (a) and (b), exactly the same selective deficit has occurred; all that differs is the pre-morbid level.

In this example, as in others we might imagine, a sizeable impairment occurs that is selective for one of our tasks, and the simple criterion will detect a dissociation, but the compound criteria will not, because there is no significant deficit. The reverse could not occur: the simple criterion could never miss a dissociation that the compound criteria finds, because the simple criterion is a sub-set of the compound criteria. This example argues that the simple criterion should be sufficient for a dissociation, and shows intuitively that statistical power could only be increased by this methodological tweak. More generally, the

chances that a differential impairment will induce a significant discrepancy must necessarily be higher than the chances that it will induce a significant discrepancy *and* a significant deficit (at least until the effect size is so large that both are always detected).

Parenthetically, Figure 1 provides a useful visualisation for understanding the strict limits on power in case-control designs. Given a task-specific impairment of 2SD (a huge effect size in most other contexts), only cases with premorbid performance at or below the mean level will be pushed below the cut-off for deficit (-1.93). Even with an unrestricted control sample size, for which the *t* distribution converges on the *z*-distribution, the one-tailed cut-off could not be higher than -1.65, so only pre-morbid performances below 0.35 would be pushed into deficit by a 2SD impairment, capping power at .64 (the area under the normal curve for $z < 0.35$). Similar considerations apply to the test of between-task difference, but in this case note that the standard deviation of the distribution of differences in controls, and thus the effect size of the discrepancy induced by a selective deficit on one task, will vary with the strength of correlation between tasks². A selective 2SD impairment on task X ($Z_{CC} = -2$) would induce a between task discrepancy with an effect size (Z_{DCC}) of 2 only if tasks X and Y correlate at .5, with the effect size being progressively reduced at lower correlations, and amplified at higher correlations.³ The test of difference will thus be increasingly sensitive as the inter-correlation between tasks increases. Accordingly, we should expect the simple criterion for dissociation to be especially useful when considering pairs of tasks with moderate to high relatedness in controls ($r > .5$). This is likely to be practically useful because, in cognitive neuropsychology, we are often most interested in dissociations between tasks that superficially seem to be related in the healthy state (e.g. recognition of living and non-living objects; processing of nouns and verbs; discrimination of pitch and timbre; etc).

² The standard deviation of the distribution of differences between correlated *z*-scores is given by $\sqrt{2 - 2r}$ where *r* is the correlation between tasks.

³The *z* score of a single-case impairment is inappropriate for testing the significance of the impairment unless the control sample is large ($n \geq 50$), but it nonetheless provides a direct standardised estimate of the effect size of the impairment. Here, I am using the notation introduced by Crawford, Garthwaite, & Porter (2010) to distinguish the effect size for the case-control test of deficit within a task (Z_{CC}) and the effect size for the case-control test of difference between tasks (Z_{DCC}).

4. Empirical simulations

To explore the consequences of the proposed change empirically, I used Monte-Carlo simulations. The general method was that, in each run, $n + 1$ cases were drawn from a bivariate normal distribution with a population-level correlation between X and Y of r , the first n cases being designated as the control sample, with case $n + 1$ as the single case of interest. I used two values of n , representing a small ($n = 10$) or a large ($n = 100$) control group, and ten values of r (from 0 to .9 in steps of .1). The X and Y values of case $n + 1$ were either used directly, to simulate a random member of the healthy population, or were perturbed to simulate a patient with known patterns of impairment on tasks X and Y (see later for details). For every combination of conditions, the simulation was run 100,000 times. The reported data summarise outcomes across runs per combination of conditions.

In each run of the simulation, the control sample statistics (mean, standard deviation, and observed correlation between X and Y) and the single case values of X and Y were used to test for a dissociation, according to compound and simple criteria. The single-case was tested first for a deficit (abnormally low score) on tasks X and Y separately, using Crawford & Howell's (1998) *modified t-test* ($p < .05$, one-tailed). The case was then tested for a discrepancy between tasks (abnormally large difference, in either direction), using Crawford & Garthwaite's (2005) *Unstandardised Difference Test* (UDT) ($p < .05$, two-tailed). The simple criterion identifies a dissociation if the test of difference is significant, and does not sub-classify the dissociation any further. The compound criteria require a significant difference *and* a significant outcome on at least one of the two tests of deficit: if one test of deficit is significant then the dissociation is classed as classical, and if both are significant then it is classed as strong. My focus is on the overall rates of dissociation, but for the compound criteria I distinguish classical and strong sub-types as an extra point of interest.

4.1. Empirical simulations: Type I errors (false positives)

To investigate Type I errors, the X and Y values of the single case ($n + 1$) were taken directly, to represent a randomly-drawn member of the healthy population. The question is how often this healthy individual would be found to show a dissociation between tasks, according to simple or compound criteria. We know that the case is in fact drawn unchanged from the healthy population, so each instance of dissociation is a Type I error, and the Type I error rate is the proportion of runs on which a dissociation is found (Figure 2).

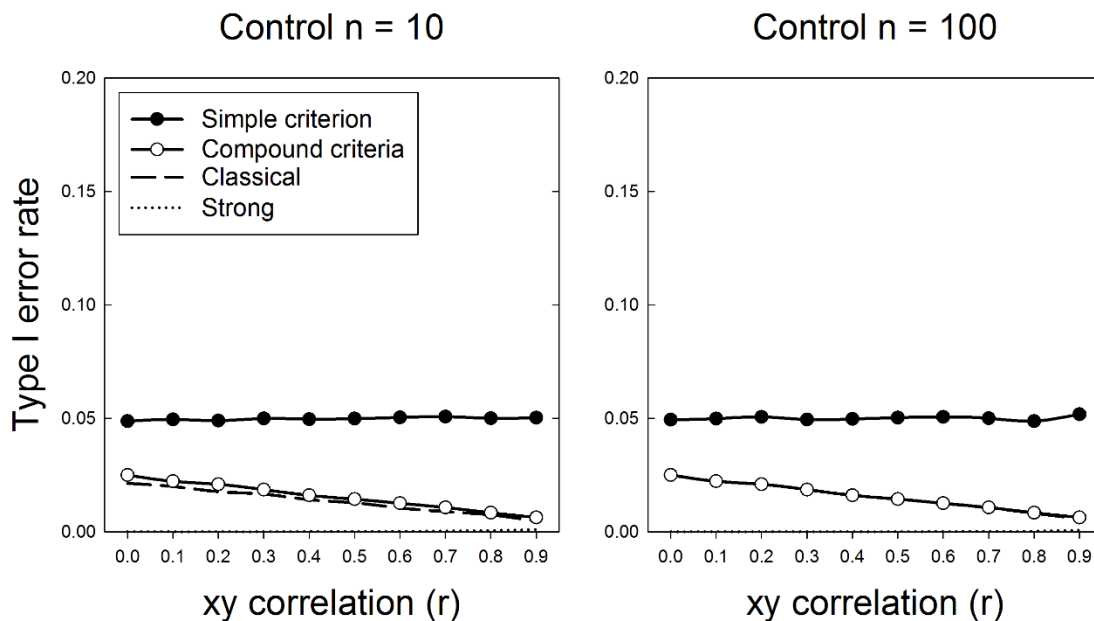


Figure 2. Type I error rate, using simple or compound criteria, for differing levels of correlation between task X and task Y ($r = 0$ to $.9$), and two different sizes of control sample ($n = 10$, $n = 100$). For the compound criteria, sub-classification of dissociations as classical or strong is indicated.

Using the simple criterion, the control of Type I error rate is essentially perfect, since it relies on a single test (UDT) optimised for case-control comparisons of difference between tasks (Crawford & Garthwaite, 2005). Unlike the simple criterion, the compound criteria do not exercise a *transparent* control of Type I error rate, which differs from the alpha criterion chosen per test, and nor is Type I error rate stable across levels of XY correlation. Because the compound criteria are contingent on a significant deficit, in addition to a between-task difference, they are more conservative than the simple criterion. This conservatism could be seen as a strength of the compound criteria, except that it must imply a corresponding reduction in power to detect true dissociations when present (i.e. increased Type II errors).

4.2. Empirical simulations: Power to detect dissociations involving a selective deficit

To investigate the power to detect dissociations involving selective deficits, Monte Carlo simulations were used, but in each run the X score of the single case ($n + 1$) was perturbed, by subtracting one, two, three or four control SDs, while the Y score was not changed. This

represents a selective deficit for task X, ranging from relatively mild to very severe. Each significant case of dissociation, in the right direction, is thus a correct detection, and the overall power, per condition, is the proportion of runs in which a dissociation was found (Figure 3).

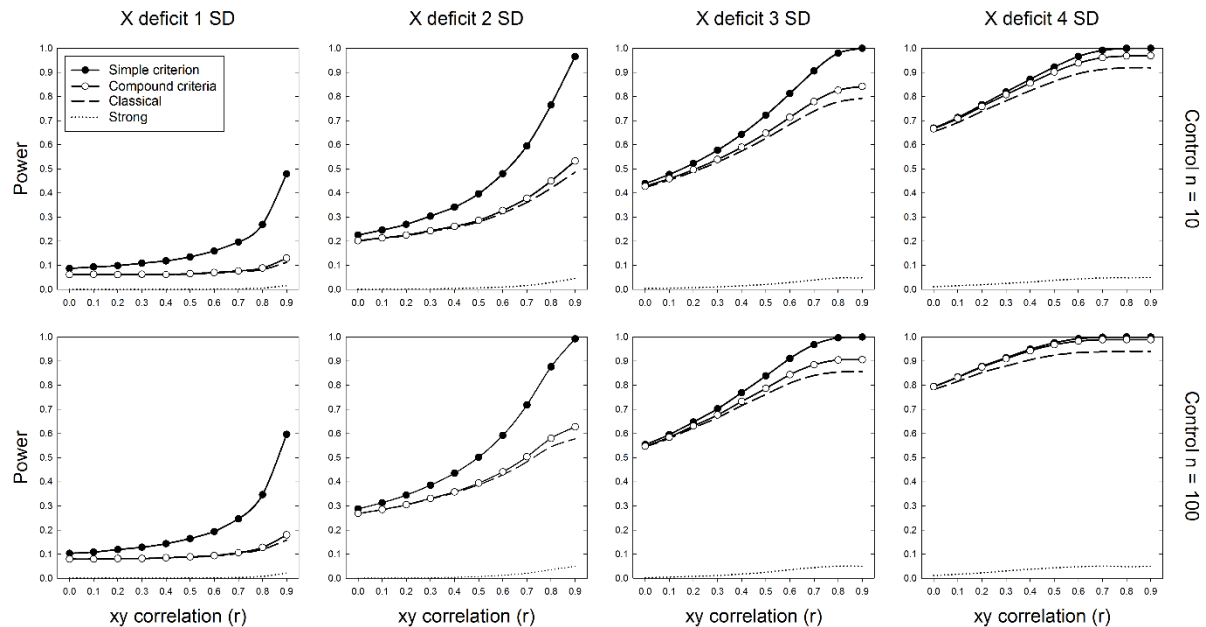


Figure 3. Power to detect dissociations, using simple or compound criteria, involving a selective deficit on task X (1-4 SD), for ten different levels of correlation between task X and task Y ($r = 0$ to $.9$), and two different sizes of control sample ($n = 10$, $n = 100$). For the compound criteria, the sub-classification of dissociations as classical or strong is shown.

Figure 3 shows that, when the deficit on X is relatively mild (1 SD), power to detect a dissociation is generally poor, and when the deficit is very severe (4 SD), power is generally good to high ($\beta > .65 - 1$), for either set of criteria. For intermediate deficits (2-3 SD), there is a modest but meaningful power advantage for the simple criterion. For instance, given a deficit on X of 2 SD, and an XY correlation of $.5$, the improvement in power of the simple criterion over the compound criteria is equivalent to that obtained by increasing the control sample size from 10 to 100. Figure 3 also confirms that the benefits of the simple criterion are negligible when X and Y are uncorrelated, but accelerate as the XY correlation increases. As noted at the end of Section 3, this is because the standard deviation of the distribution of difference scores reduces with increasing correlation between tasks, so that a given size of

single-task deficit is increasingly amplified in terms of the effect size of the between-task difference (Z_{DCC}).

4.3. Empirical simulations: Power to detect dissociations involving differential deficits

Finally, to investigate power to detect dissociations involving differential deficits, in each run of the simulation, the X score of the single case ($n + 1$) was perturbed by subtracting three, four or five control SDs, and the Y score was simultaneously perturbed to a lesser degree, by subtracting one, two or three SDs respectively. Each case of dissociation, in the right direction, was considered a correct detection, and the overall power per condition is the proportion of runs in which a dissociation was found (Figure 4).

In this case, the power to detect a dissociation is hardly affected by the criterion used. This equivalence arises because the major (X) deficit is always in the very large range (3-5 SD), so is almost always identified by the test of deficit, with the consequence that the compound criteria detect a dissociation whenever the test of difference is also significant. In effect, the finding of dissociation by the compound criteria becomes contingent on the same test of difference that constitutes the simple criterion. However, whether the compound criteria classify the dissociation as ‘strong’ or ‘classical’ depends heavily on the power to detect the more minor (Y) deficit. If the Y deficit is relatively mild (1 SD), the dissociation is usually misclassified as classical. Only when the Y deficit is itself very large (3 SD) is the dissociation reliably classified correctly as strong. These problems of misclassification by the compound criteria have been demonstrated and discussed before (Crawford & Garthwaite, 2006a).

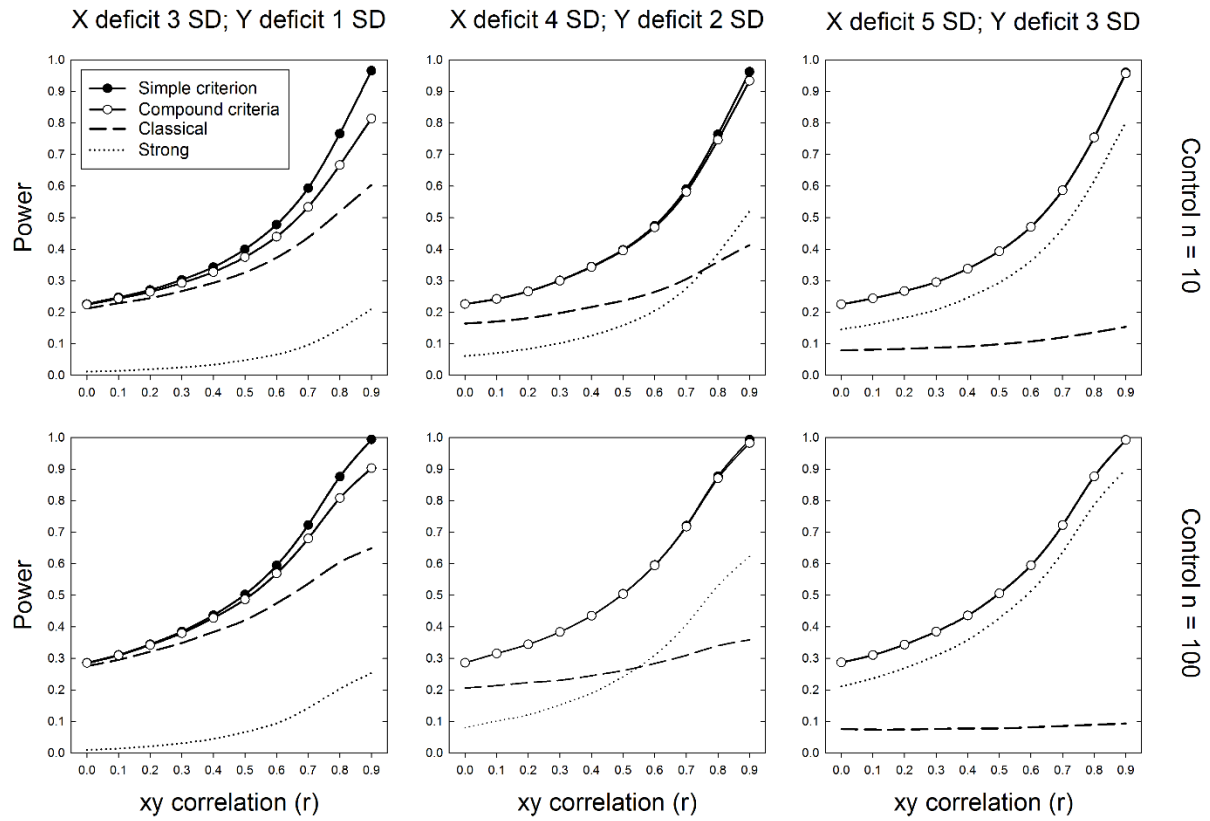


Figure 4. Power to detect dissociations, using simple or compound criteria, involving a differential deficit between task X (3-5 SD) and task Y (1-3 SD), for ten different levels of correlation between task X and task Y ($r = 0$ to $.9$), and two different sizes of control sample ($n = 10$, $n = 100$). For the compound criteria, the sub-classification of dissociations as classical or strong is shown.

5. General discussion

The empirical simulations show that simplifying the criteria for neuropsychological dissociation, by collapsing to the requirement for a significantly extreme difference between tasks, increases power to detect dissociations, with transparent and stable control of Type I error. The increases in power are far from negligible, and apply across a practically relevant range of conditions. These differences all stem from the relaxation of a traditional requirement that has been baked into the accepted, compound criteria for dissociation: namely, that a patient must display a deficit for at least one function before a dissociation from some other function can be considered. Prior to the advent of appropriate statistical tools for case-control comparisons, neuropsychological deficits were typically stark qualitative contrasts, and a deficit in one function was indeed a precondition for dissociation, if only because a patient with no evident deficits would have no cause to be singled out for study. However, statistical tests have now been developed that allow for a positive test of abnormal difference between two tasks. In this contemporary context, the requirement for a deficit, to confirm a dissociation, is anachronistic and unnecessarily limiting.

The power benefits of the simple criterion accrue from those occasions in which a between-task difference is detected but a within-task deficit is not. It is worth rehearsing some of the reasons that this may occur. As was illustrated in Section 3 (Figure 1), the higher the pre-morbid ability for a given task, the less visible any acquired deficit will be. This is well-known to clinicians, who must make diagnostic judgements. A standard cognitive exam is less likely, on average, to disclose early signs of dementia in a University Professor than in a University Janitor. Conversely, a close friend or family member may clearly see changes in a loved-one, though that person is well-within statistical limits of normality. We can strive to improve diagnostic accuracy through stratified or multivariate standardisation, adjusting for relevant variables such as age, sex, and years of education (e.g. Capitani, 1997; Capitani & Laiacina, 1997); but this requires large-scale data collection from the relevant culture, and would be impractical for all but the most widely-used neuropsychological tests. Moreover, even a closely-matched control sample is at best an approximate proxy for the premorbid capacities of any particular patient. This is reflected in the clinical wisdom that, when examining a patient across a battery of tests, one should not consider each test in isolation, but pay careful attention to the pattern across tests, being alert to scores that seem inconsistent with the patient's overall profile. The between-task discrepancies may disclose a differential deficit, even when the within-task performances are unremarkable.

The power to detect a differential deficit will increase with the inter-task correlation, as shown graphically by the diverging power lines in Figure 2. This is of major practical relevance for researchers, because dissociations are of little theoretical interest between functions that seem unrelated *a priori* (e.g. noun processing and motor imagery). We are more often seeking dissociations between putative functions that seem related in the healthy state (e.g. noun and verb processing; motor imagery for hands and feet). A prolific journalist who suffers a selective impairment for nouns may still score within normal limits on a lexical decision task for nouns, due to her premorbid facility with words; but the contrast with her excellent (preserved) performance for verbs could be dramatic. In this example, verb processing would act somewhat like a pre-morbid benchmark (or covariate) to reveal the differential deficit for nouns.

In many research settings, inter-correlations between tasks will be raised by a host of less specific factors, because the tasks compared are usually matched quite closely, except for critical features. This means that individual differences (e.g. sensory acuity; psychomotor reaction time; attention and motivation, general intelligence), which increase the spread of normal performance and limit power to detect deficits within a task, will strengthen the correlation between tasks and increase power to detect a differential deficit. Moreover, multiple measures from a single task may be separable in principle but correlated in practice. Bundesen's (1990) analysis of the whole report task, which has been widely applied in clinical settings (Habekost & Starrfelt, 2009), gives independent measures of processing speed (C) and visual working memory capacity (k). Studies of healthy participants indicate that these parameters correlate, even as highly as $r = .72$ (Habekost, Petersen, & Vangkilde, 2014). Accordingly, the power to detect a theoretically interesting dissociation between C and k, would be enhanced using the simple criterion.

My own file drawer provides another illustration, whilst highlighting a further measurement issue that can limit our ability to detect frank deficits. Using a laterality discrimination task for pictures of effectors (a widely used test of motor imagery, after Parsons, 1987), I found that the performance of 12 older adults ranged from near-chance (56%) to near-perfect (99%), but that the correlation between parallel versions of the task for hands and feet was very tight ($r = .82$) and the mean difference correspondingly low (2%). If we imagine recruiting a single-case with a central or peripheral abnormality of the upper-limbs (e.g. lesion in primary motor cortex, dystonia, amputation), then the scope for a frank impairment of motor imagery for hands is slim, because the control distribution encroaches

on the worst possible performance; yet if the accuracy for hands were just 20% lower than that for feet in our single case, the differential deficit would be significant ($p < .05$, two-tailed). Considering that the control range on some tasks leaves relatively little scope for detection of impaired performance, the benefits of the simple criterion may be even greater in practice than the present empirical simulations suggest.

So, given these evident benefits, is there any reason *not* to simplify the criterion for dissociation to a single test of differential deficit? One objection might be that we would lose the ability to discriminate classical and strong forms of dissociation, which support different strengths of inference. This objection would be ill-founded. Taking the test of difference as a direct test of dissociation does not preclude testing for frank deficits on each task individually. Arguably, we would gain a category: the simple dissociation (without frank deficits) to add to the classical dissociation (with one frank deficit) and the strong dissociation (with two frank deficits). However, it is debatable how useful such categories actually are. As demonstrated in Section 4.3, strong dissociations are often mistaken for classical, and since we cannot formally confirm preserved performance on the ‘unimpaired’ task, the classical dissociation is only ever “putative” (Crawford & Garthwaite, 2006). Similarly, for the simple dissociation, we cannot formally confirm the absence of deficits on either task; indeed the point of the differential deficit is that we infer that (at least) one task has been compromised. Positive findings of frank deficits within tasks may allow us to say more with confidence about the patient’s condition; but the top of this informational hierarchy (dissociation with two frank deficits) does not give us greater *inferential* power, just greater certainty that the deficit is *not* selective to a single task. The demonstration of a selective deficit is always more-or-less tentative.

These considerations suggest that we should give less weight to the inherently uncertain classification of dissociation sub-types, and focus on the between-task difference. Some studies have already used this approach, inferring dissociations directly from a discrepancy between tasks (e.g. Toraldo, Laiacona, Pagani, Mandelli, & Capitani, 2014); but it is likely that many more dissociations of potential interest have gone unreported, due to a lack of within-task deficits. The simple criterion reduces the importance of within-task outcomes, and may steer us away from a categorical contrast between classical and strong dissociation types. To characterise the strength of dissociation, we might think less categorically, and instead use a continuous index of the degree of separation between tasks. The obvious candidate would be the effect size for the test of difference (Z_{DCC}), as calculated

from the difference between the patient's standardised (z) scores on the two tasks, divided by the standard deviation of the difference between two correlated z scores (Crawford et al., 2010). Additional testing for deficits in the individual tasks would provide further information about the patient's condition, but would not define the dissociation. However, before adopting an effect-size index to guide theorising, we would need to give more study to the biasing effects that various factors could have upon its estimation, including control sample size, the effects of non-normality, and the selective reporting of positive dissociations: a case-control variant of the so-called 'winner's curse', whereby selective reporting of positive outcomes biases estimated effects sizes upward, especially for poorly-powered studies (Button et al., 2013; Ioannidis, 2008).

A second, related objection might be that a simple dissociation is only a relative difference, so without frank deficits to anchor our conclusions, our result would be ambiguous. We might have discovered an impairment on task X, or an *improvement* on task Y. This is only a problem in principle, however. In practice, we could make a default assumption that a relative difference, when otherwise ambiguous, reflects a deficit for the poorer task, consistent with the idea that tests of deficit should be one-tailed, because "*Situations in which improvement in performance over premorbid levels occurs as the result of brain damage are highly specific and very rare*" (Crawford et al., 2003, p359). It is also worth noting that, in contrast to the one-tailed tests of deficit, the compound criteria specify a two-tailed test of difference (Crawford & Garthwaite, 2005, 2007b; Crawford et al., 2003). This is because, before examining the data, we have no prior commitment to the direction of dissociation. But this implies a relatively exploratory approach to dissociations. In practice, we may often have a strong prediction about the direction of dissociation, based on knowledge of a patient's lesion, and/or their behaviour in other circumstances, which might license a one-tailed test of difference. As ever, the choice of an appropriate alpha, and its directional allocation, should be flexible to fit the context, as long as it is not post-hoc.

A third objection might be that the simple criterion would be too liberal. Control samples are often limited, and assumptions of normality may not hold, or may not be robustly checkable in small samples, so we should prefer to be conservative. There could thus be pragmatic value to having to pass an additional test of deficit to confirm a dissociation. The concern here is valid, but retaining the compound criteria as a 'belt-and-braces' solution is not an elegant way to address it. The simulations in Section 4.1 show that, if assumptions of normality hold, the simple criterion gives robust and transparent control of Type I error. We

should aim to collect well-distributed data, from reasonable control sample sizes, if at all possible. However, if the normality of test performance is in doubt, researchers could explicitly set a more stringent alpha, as recommended by Crawford & Garthwaite (2006). The simple criterion gives us direct control over Type I error rate, so it will be as conservative or liberal as we choose to make it.

Conclusion

The conventional criteria for a neuropsychological dissociation demand that a patient has a frank deficit on at least one task, as well as a significant between-task difference, relative to controls. I have argued that the requirement for positive evidence of a within-task deficit should be relaxed, so that the criterion for a dissociation simplifies to an appropriate test of the between-task difference. This modest methodological tweak increases power to detect true dissociations across a range of practically-relevant conditions, whilst providing excellent control over Type I error. Giving primacy to the test of difference may encourage a less categorical approach to evaluating dissociations, so that the degree of dissociation, rather than the sub-division into unreliable subtypes, can inform theorising. Of course, dissociations never permit unambiguous conclusions in and of themselves, and the statistical outcome of a simple test of difference will always be part of a bigger picture, informed by the overall pattern of a patient's performance, available information about pre-morbid abilities, and the wider empirical and theoretical context (cf. McShane, Gal, Gelman, Robert, & Tackett, 2017). Dissociations nonetheless offer important evidence to constrain and guide models of normal and damaged brain function, and the proposed simplification may further enhance the power and utility of this time-honoured approach.

Acknowledgements

I am grateful to Alessio Toraldo, Anna Sedda, Ratko Radakovic, Sergio Della Sala, and Tim Shallice for critical comments on an earlier draft of this manuscript.

References

- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, 97(4), 523–547.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience*, 14(5), 365–76.
- Capitani, E. (1997). Normative Data and Neuropsychological Assessment. Common Problems in Clinical Practice and Research. *Neuropsychological Rehabilitation*, 7(4), 295–310.
- Capitani, E., & Laiacona, M. (1997). Composite neuropsychological batteries and demographic correction: Standardization based on equivalent scores, with a review of Published Data. *Journal of Clinical and Experimental Neuropsychology*, 19(6), 795–809.
- Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. *Brain and Cognition*, 5(1), 41–66.
- Caramazza, A., & McCloskey, M. (1988). The case for single-patient studies. *Cognitive Neuropsychology*, 5(5), 517–527.
- Coltheart, M. (2001). Assumptions and methods in cognitive neuropsychology. In B. Rapp (Ed.), *The handbook of cognitive neuropsychology: What deficits reveal about the human mind* (pp. 3–21). New York: Psychology Press.
- Coltheart, M. (2017). The assumptions of cognitive neuropsychology: Reflections on Caramazza (1984, 1986). *Cognitive Neuropsychology*, 34(7–8), 397–402.
- Crawford, J. R., & Garthwaite, P. H. (2005). Testing for Suspected Impairments and Dissociations in Single-Case Studies in Neuropsychology: Evaluation of Alternatives Using Monte Carlo Simulations and Revised Tests for Dissociations. *Neuropsychology*, 19(3), 318–331.
- Crawford, J. R., & Garthwaite, P. H. (2006a). Detecting dissociations in single-case studies: Type I errors, statistical power and the classical versus strong distinction. *Neuropsychologia*, 44(12), 2249–2258.
- Crawford, J. R., & Garthwaite, P. H. (2006b). Methods of testing for a deficit in single-case studies: Evaluation of statistical power by Monte Carlo simulation. *Cognitive Neuropsychology*, 23(6), 877–904.
- Crawford, J. R., & Garthwaite, P. H. (2007a). Comparison of a single case to a control or normative sample in neuropsychology: Development of a Bayesian approach. *Cognitive*

- Neuropsychology*, 24(768491913), 343–372.
- Crawford, J. R., & Garthwaite, P. H. (2007b). Comparison of a single case to a control or normative sample in neuropsychology: Development of a Bayesian approach. *Cognitive Neuropsychology*, 24(768491913), 343–372.
- Crawford, J. R., Garthwaite, P. H., Azzalini, A., Howell, D. C., & Laws, K. R. (2006). Testing for a deficit in single-case studies: Effects of departures from normality. *Neuropsychologia*, 44(4), 666–677.
- Crawford, J. R., Garthwaite, P. H., & Gray, C. D. (2003). Wanted: fully operational definitions of dissociations in single-case studies. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 39(2), 357–370.
- Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, 27(3), 245–260.
- Crawford, J. R., Garthwaite, P. H., & Ryan, K. (2011). Comparing a single case to a control sample: Testing for neuropsychological deficits and dissociations in the presence of covariates. *Cortex*, 47(10), 1166–1178.
- Crawford, J. R., & Howell, D. C. (1998). Comparing an Individual's Test Score Against Norms Derived from Small Samples. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D)*, 12(4), 482–486.
- Crawford, J. R., Howell, D. C., & Garthwaite, P. H. (1998). Payne and Jones Revisited: Estimating the Abnormality of Test Score Differences Using a Modified Paired Samples t Test. *Journal of Clinical and Experimental Neuropsychology*, 20(6), 898–905.
- Dunn, J. C., & Kirsner, K. (2003). What can we infer from double dissociations? *Cortex*, 39(1), 1–7.
- Ellis, A. W., & Young, A. W. (1988). *Human Cognitive Neuropsychology*. Hove, UK.: Erlbaum.
- Habekost, T., Petersen, A., & Vangkilde, S. (2014). Testing attention: Comparing the ANT with TVA-based assessment. *Behavior Research Methods*, 46(1), 81–94.
- Habekost, T., & Starrfelt, R. (2009). Visual attention capacity: A review of TVA-based patient studies: Cognition and Neurosciences. *Scandinavian Journal of Psychology*, 50(1), 23–32.
- Ioannidis, J. P. A. (2008). Why Most Discovered True Associations Are Inflated. *Epidemiology*, 19(5), 640–648.

- McIntosh, R. D., & Brooks, J. L. J. L. J. L. J. L. (2011). Current tests and trends in single-case neuropsychology. *Cortex*, 47(10), 1151–1159.
- McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2017). Abandon Statistical Significance. *arXiv Preprint*.
- Parsons, L. M. (1987). Imagined spatial transformations of one's hands and feet. *Cognitive Psychology*, 19(2), 178–241.
- Patterson, K., & Plaut, D. C. (2009). “Shallow Draughts Intoxicate the Brain”: Lessons from Cognitive Science for Cognitive Neuropsychology. *Topics in Cognitive Science*, 1(1), 39–58.
- Shallice, T. (1979). Case study approach in neuropsychological research. *Journal of Clinical Neuropsychology*, 1(3), 183–211.
- Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge University Press.
- Shallice, T. (2015). Cognitive neuropsychology and its vicissitudes: The fate of Caramazza's axioms. *Cognitive Neuropsychology*, 32(7–8), 385–411.
- Toraldo, A., Laiacona, M., Pagani, R., Mandelli, A., & Capitani, E. (2014). Perceptual and response-related components of unilateral neglect may evolve independently of one another: Evidence from five single-case studies. *Neurocase*, 20(3), 241–259.